

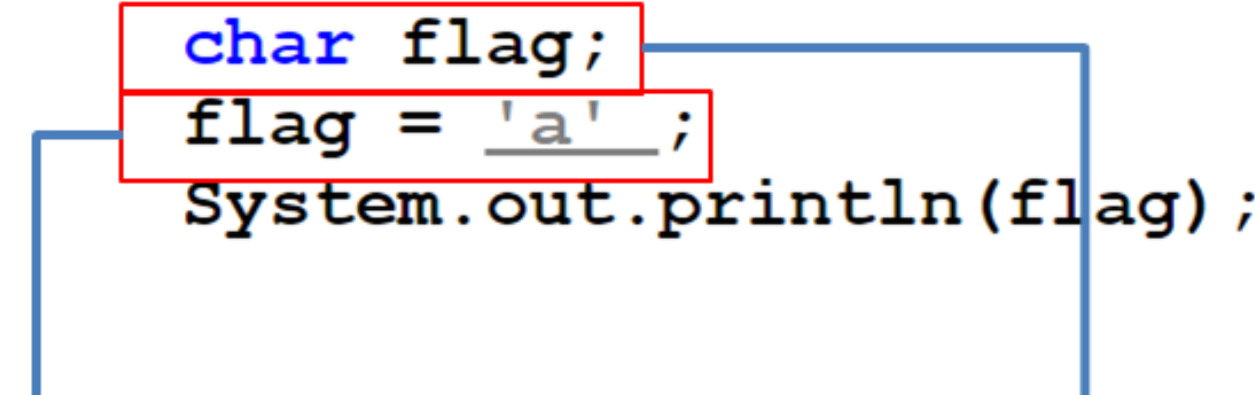
Java for Beginners

CHARACTER AND ASCII

ERIC Y. CHOU, PH.D.

IEEE SENIOR MEMBER

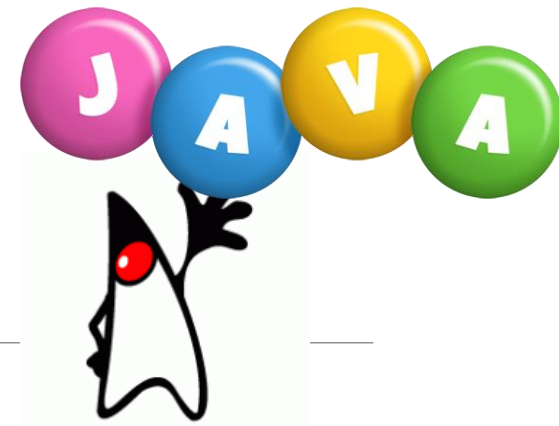
```
// A Java program to demonstrate char data type
class CharacterExample1 {
    public static void main(String args[])
    {
        char flag;
        flag = 'a';
        System.out.println(flag);
    }
}
```



The diagram illustrates the two parts of the variable 'flag' in the provided Java code. A red box highlights the declaration 'char flag;' and the initialization 'flag = 'a';'. A blue arrow originates from the 'char' keyword in the declaration and points to a box labeled 'char variable Declaration'. Another blue arrow originates from the 'flag = 'a';' line and points to a box labeled 'char variable Initialization'.

char variable Initialization

char variable Declaration



Java char Data Type

```
char ch = 'A';
```

```
char chou = '\u5488';
```

```
char chou = 0x5468;
```

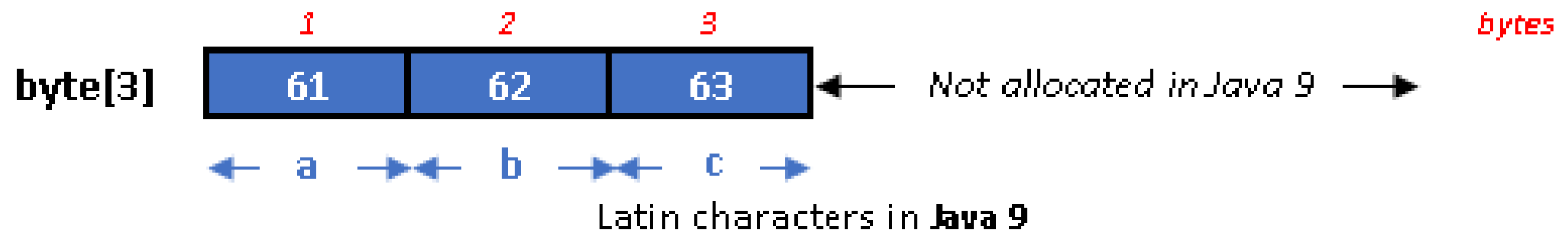
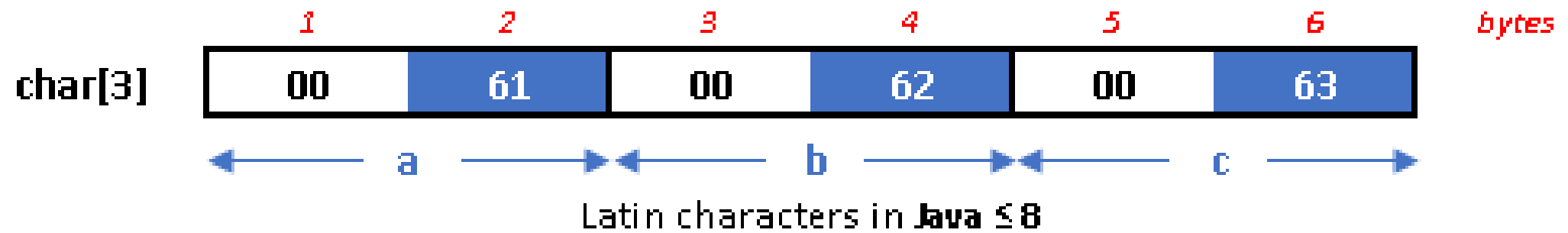
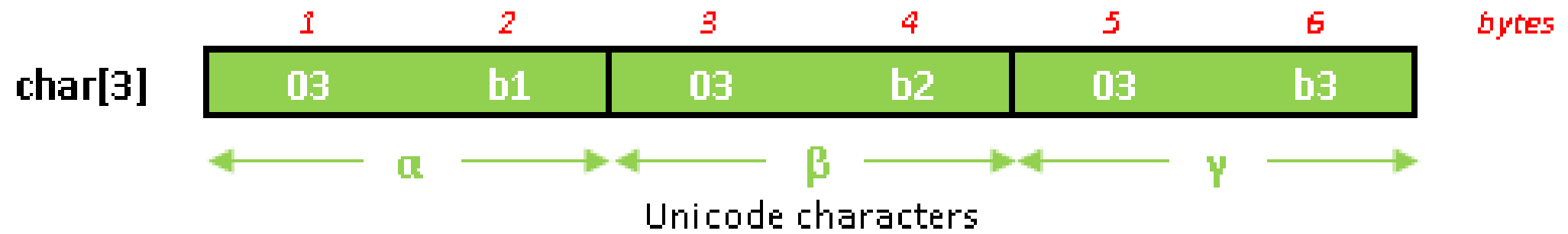
```
String str = "A";
```

```
String chou = "\u5488";
```

```
String chou = new String("\u5488")
```

Unsigned Character Arithmetic:

```
char chB = (char) (ch + 1); // 'B'
```



declare a variable (object name)

invoke a constructor to create an object

`String s;`

`s = new String("Hello, World");`

`char c = s.charAt(4);`

object name

*invoke an instance method
that operates on the object's value*

ASCII

	00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F
00	<u>NUL</u> 0000	<u>STX</u> 0001	<u>SOT</u> 0002	<u>ETX</u> 0003	<u>EOT</u> 0004	<u>ENQ</u> 0005	<u>ACK</u> 0006	<u>BEL</u> 0007	<u>BS</u> 0008	<u>HT</u> 0009	<u>LF</u> 000A	<u>VT</u> 000B	<u>FF</u> 000C	<u>CR</u> 000D	<u>SO</u> 000E	<u>SI</u> 000F
10	<u>DLE</u> 0010	<u>DC1</u> 0011	<u>DC2</u> 0012	<u>DC3</u> 0013	<u>DC4</u> 0014	<u>NAK</u> 0015	<u>SYN</u> 0016	<u>ETB</u> 0017	<u>CAN</u> 0018	<u>EM</u> 0019	<u>SUB</u> 001A	<u>ESC</u> 001B	<u>FS</u> 001C	<u>GS</u> 001D	<u>RS</u> 001E	<u>US</u> 001F
20	<u>SP</u> 0020	!	"	#	\$	%	&	'	()	*	+	,	-	.	/
30	0 0030	1 0031	2 0032	3 0033	4 0034	5 0035	6 0036	7 0037	8 0038	9 0039	:	;	<	=	>	?
40	@ 0040	A 0041	B 0042	C 0043	D 0044	E 0045	F 0046	G 0047	H 0048	I 0049	J 004A	K 004B	L 004C	M 004D	N 004E	O 004F
50	P 0050	Q 0051	R 0052	S 0053	T 0054	U 0055	V 0056	W 0057	X 0058	Y 0059	Z 005A	[005B	\ 005C] 005D	^ 005E	_ 005F
60	` 0060	a 0061	b 0062	c 0063	d 0064	e 0065	f 0066	g 0067	h 0068	i 0069	j 006A	k 006B	l 006C	m 006D	n 006E	o 006F
70	p 0070	q 0071	r 0072	s 0073	t 0074	u 0075	v 0076	w 0077	x 0078	y 0079	z 007A	{ 007B	 007C	} 007D	~ 007E	<u>DEL</u> 007F
80	Ç 00C7	ü 00FC	é 00E9	â 00E2	ä 00E4	à 00E0	å 00E5	ç 00E7	ê 00EA	ë 00EB	è 00E8	ì 00EF	í 00EE	î 00EC	Ä 00C4	Å 00C5
90	É 00C9	æ 00E6	Æ 00C6	ø 00F4	ö 00F6	ò 00F2	û 00FB	ù 00F9	ÿ 00FF	Ö 00D6	Ü 00DC	¢ 00A2	£ 00A3	¥ 00A5	€ 20A7	f 0192
A0	á 00E1	í 00ED	ó 00F3	ú 00FA	ñ 00F1	Ñ 00D1	ª 00AA	º 00BA	¿ 00BF	¬ 2310	¬ 00AC	½ 00BD	¾ 00BC	¡ 00A1	« 00AB	» 00BB
B0	▯ 2591	▯ 2592	▯ 2593	 2502	 2524	 2561	 2562	 2556	 2555	 2563	 2551	 2557	 255D	 255C	 255B	 2510
C0	L 2514	┐ 2534	┐ 252C	┐ 251C	— 2500	┐ 253C	┐ 255E	┐ 255F	┐ 255A	┐ 2554	┐ 2569	┐ 2566	┐ 2560	= 2550	┐ 256C	┐ 2567
D0	┐ 2568	┐ 2564	┐ 2565	┐ 2559	┐ 2558	┐ 2552	┐ 2553	┐ 256B	┐ 256A	┐ 2518	┐ 250C	▯ 2588	▯ 2584	▯ 258C	▯ 2590	▯ 2580
E0	α 03B1	β 00DF	Γ 0393	π 03C0	Σ 03A3	σ 03C3	μ 00B5	τ 03C4	Φ 03A6	Θ 0398	Ω 03A9	δ 03B4	∞ 221E	φ 03C6	ε 03B5	∩ 2229
F0	≡ 2261	± 00B1	≥ 2265	≤ 2264	∫ 2320	∫ 2321	÷ 00F7	≈ 2248	° 00B0	· 2219	· 00B7	√ 221A	² 207F	² 00B2	■ 25A0	<u>NBSP</u> 00A0

Unicode

00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F
10	11	12	13	14	15	16	17	18	19	1A	1B	1C	1D	1E	1F
20	21	22	23	24	25	26	27	28	29	2A	2B	2C	2D	2E	2F
30	31	32	33	34	35	36	37	38	39	3A	3B	3C	3D	3E	3F
40	41	42	43	44	45	46	47	48	49	4A	4B	4C	4D	4E	4F
50	51	52	53	54	55	56	57	58	59	5A	5B	5C	5D	5E	5F
60	61	62	63	64	65	66	67	68	69	6A	6B	6C	6D	6E	6F
70	71	72	73	74	75	76	77	78	79	7A	7B	7C	7D	7E	7F
80	81	82	83	84	85	86	87	88	89	8A	8B	8C	8D	8E	8F
90	91	92	93	94	95	96	97	98	99	9A	9B	9C	9D	9E	9F
A0	A1	A2	A3	A4	A5	A6	A7	A8	A9	AA	AB	AC	AD	AE	AF
B0	B1	B2	B3	B4	B5	B6	B7	B8	B9	BA	BB	BC	BD	BE	BF
C0	C1	C2	C3	C4	C5	C6	C7	C8	C9	CA	CB	CC	CD	CE	CF
D0	D1	D2	D3	D4	D5	D6	D7	D8	D9	DA	DB	DC	DD	DE	DF
E0	E1	E2	E3	E4	E5	E6	E7	E8	E9	EA	EB	EC	ED	EE	EF
F0	F1	F2	F3	F4	F5	F6	F7	F8	F9	FA	FB	FC	FD	FE	FF

- Latin script
- Non-Latin European scripts
- African scripts
- Middle Eastern and Southwest Asian scripts
- South and Central Asian scripts
- Southeast Asian scripts
- East Asian scripts
- CJK characters
- Indonesian and Oceanic scripts
- American scripts
- Notational systems
- Symbols
- Private use
- UTF-16 surrogates
- Unallocated code points

As of Unicode 10.0

A	Ω	語	𠂇
00000041	000003A9	00008A9E	00010384

UTF-32

A	Ω	語	𠂇
0041	03A9	8A9E	D800 DF84

UTF-16

A	Ω	語	𠂇
41	CE A9	E8 AA 9E	F0 90 8E 84

UTF-8

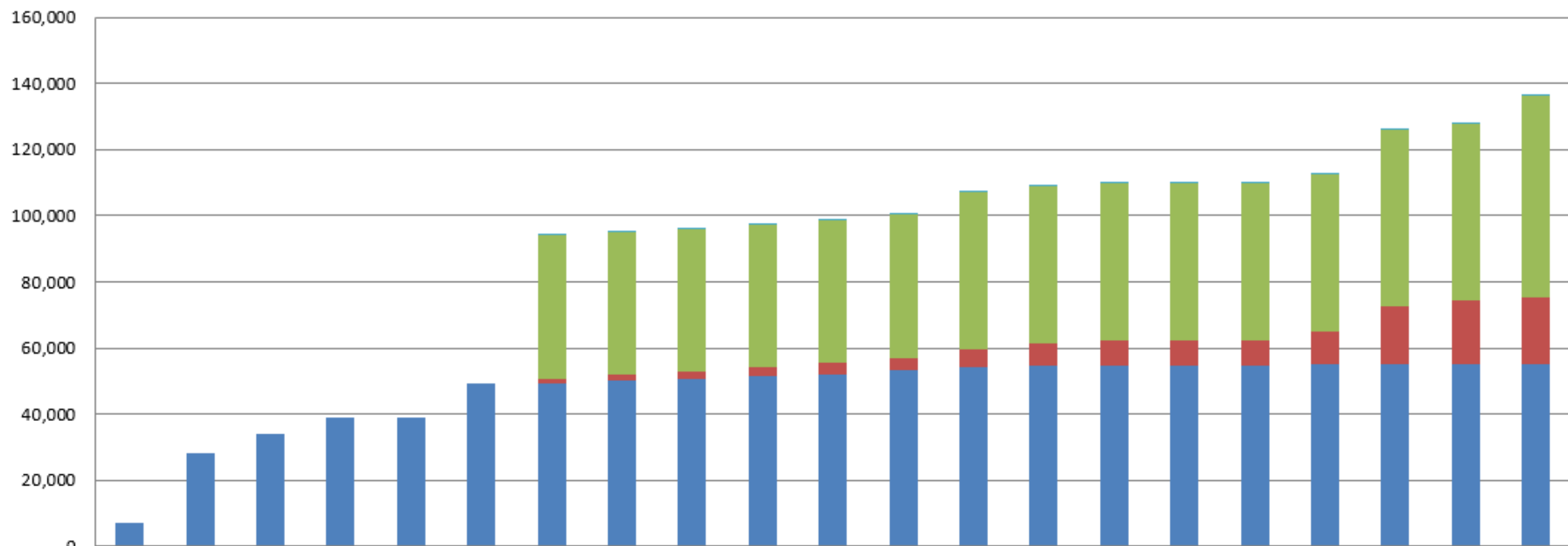
Character	UTF-16	UTF-8	UCS-2
A	0041	41	0041
c	0063	63	0063
Ö	00F6	C3 B6	00F6
𠂇	4E9C	E4 BA 9C	4E9C
♫	D834 DD1E	F0 9D 84 9E	N/A

character	encoding	bits
A	UTF-8	01000001
A	UTF-16	00000000 01000001
A	UTF-32	00000000 00000000 00000000 01000001
あ	UTF-8	11100011 10000001 10000010
あ	UTF-16	00110000 01000010
あ	UTF-32	00000000 00000000 00110000 01000010

Internal encodings

- C strings are byte arrays
- C++ bytes, UTF-16, or UTF-32
- Java UTF-16
- .NET UTF-16
- Python UTF-16
- Ruby ???
- JavaScript UTF-16 or UCS-2 (poorly defined)
- PHP strings are byte arrays¹⁾
- Perl UTF-8

Total Unicode Characters in each Version



	1991	1992	1993	1996	1998	1999	2001	2002	2003	2005	2006	2008	2009	2010	2012	2012	2013	2014	2015	2016	2017
	1.0.0	1.0.1	1.1	2.0	2.1	3.0	3.1	3.2	4.0	4.1	5.0	5.1	5.2	6.0	6.1	6.2	6.3	7.0	8.0	9.0	10.0
	7,129	28,327	34,168	38,885	38,887	49,194	94,140	95,156	96,382	97,655	99,024	100,648	107,296	109,384	110,116	110,117	110,122	112,956	126,434	128,172	136,690
Plane 12 (SSP)	0	0	0	0	0	0	97	97	337	337	337	337	337	337	337	337	337	337	337	337	337
Plane 3 (TIP)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Plane 2 (SIP)	0	0	0	0	0	0	43,253	43,253	43,253	43,253	43,253	43,253	47,402	47,624	47,624	47,624	47,624	47,624	53,386	53,386	60,859
Plane 1 (SMP)	0	0	0	0	0	0	1,594	1,594	2,128	2,390	3,419	3,727	5,258	6,893	7,497	7,497	7,497	10,004	17,595	19,277	20,265
Plane 0 (BMP)	7,129	28,327	34,168	38,885	38,887	49,194	49,196	50,212	50,664	51,675	52,015	53,331	54,299	54,530	54,658	54,659	54,664	54,991	55,116	55,172	55,229